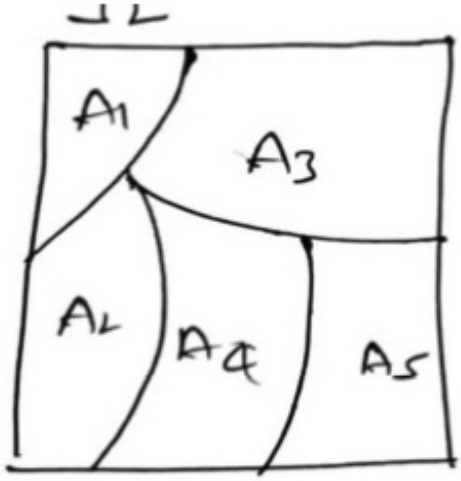
Probability Part of Sigmoid Neuron

medium.com/@manveetdn/notes-on-basic-probability-part-of-sigmoid-neuron-padhai-onefourthlabs-course-a-first-course-on-c6a4038b072e

> **Disclaimer: This is notes on "*Probability Part of Sigmoid Neuron*" Lesson (PadhAI onefourthlabs course "A First Course on Deep Learning")**

Denoting Universal and individual sets.

1. For any event A P(A ≥ 0 always probability any event lies between 0 and 1.
2. If **A1,A2,A3,.....An** are disjoint (dis joint means **Ai∩Aj = Ø ∀i≠j**)
3. **Union of all i events probability** = **Sum of all individual probabilities of i events.**
4. If **Ω** is the universal set (set containing all sets)
5. **P(Ω) = 1** (Probability of universal set)



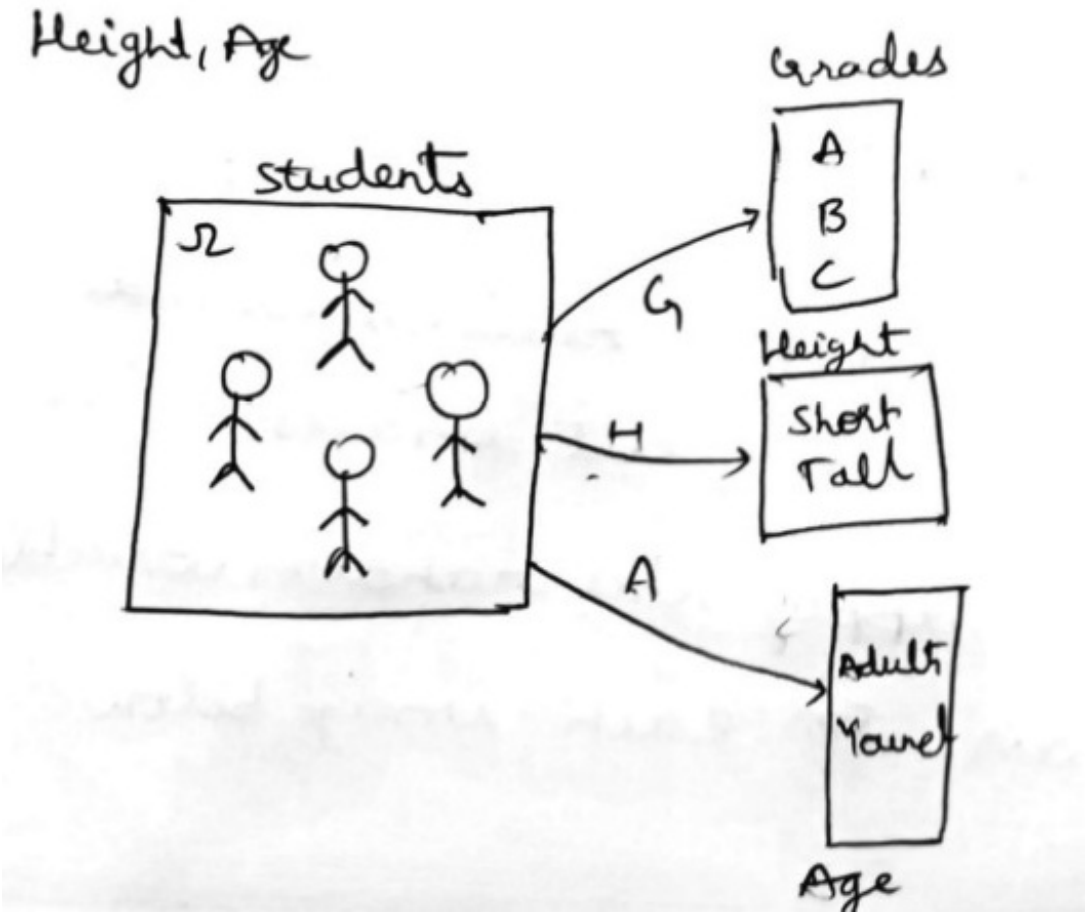
Example :

Suppose a student gets one S possible grades in a course : **A,B,C**

P(grade = A),P(grade = B) ,P(grade = C)

Actually in this case normally probabilities is calculated as (Numbers of 'A' Grades)/(Total number of students)

Actually coming to students we can have **multiple properties** associated with like **Grades, Height, Age**.



Each multiple-properties related to each student.

Random Variable:

A random variable is a function which maps each outcomes **Ω** to a value**.**

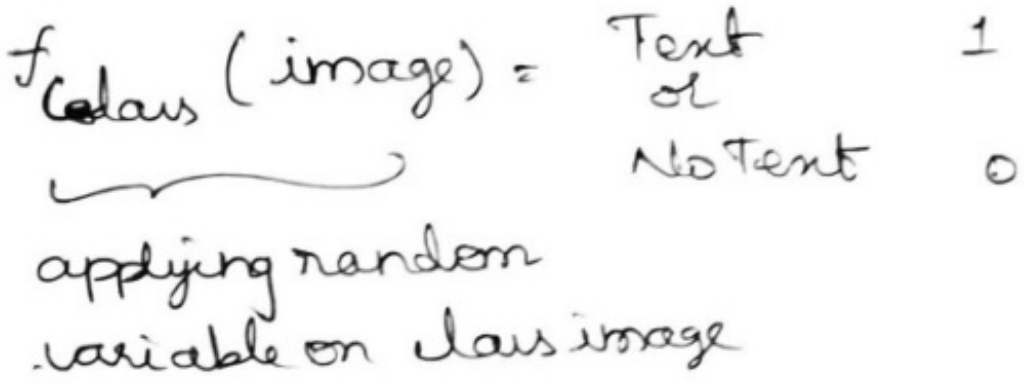
I the students example G(or fgrade) maps each student in **Ω** to a value: **A,B or C.**

The event Grade = A is a shorthand fr the event **{w ∈ Ω: Fgrade=A }.**

Actually here we can take capstone project as example is that we ave a heap of images and we need to decide whether it has text or no text

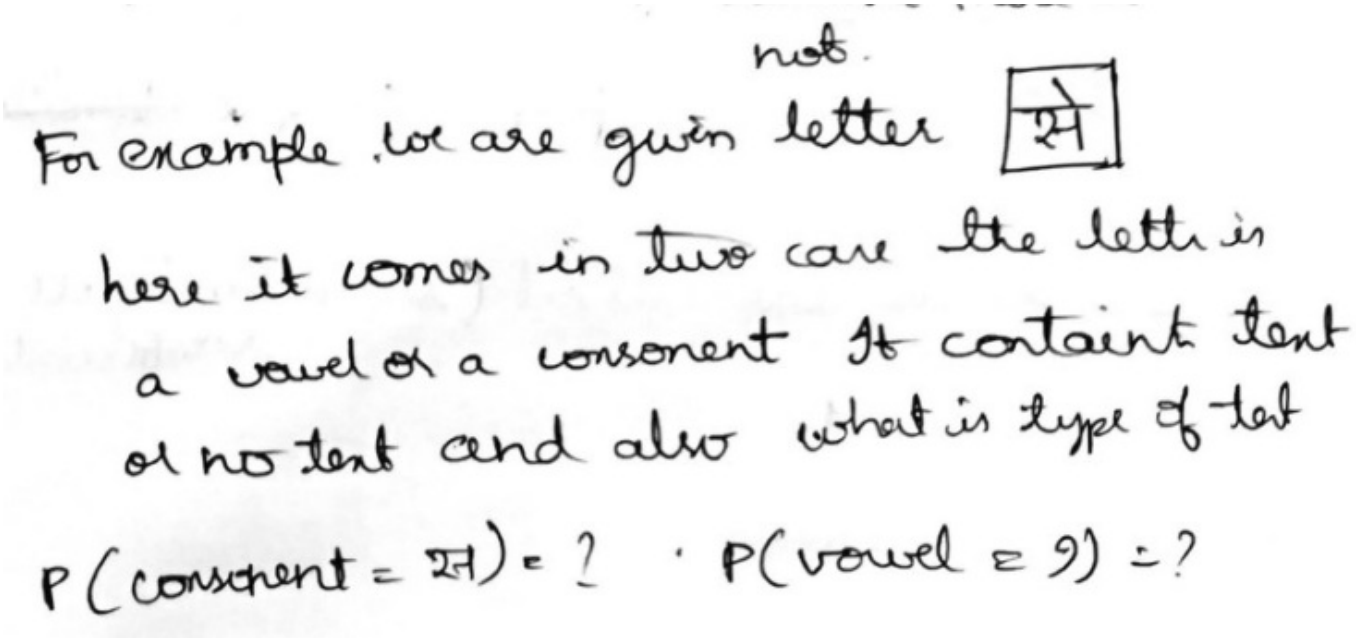
Therefore,We will initialise a folder having all images.

And if we apply this random variable called class on each image below.

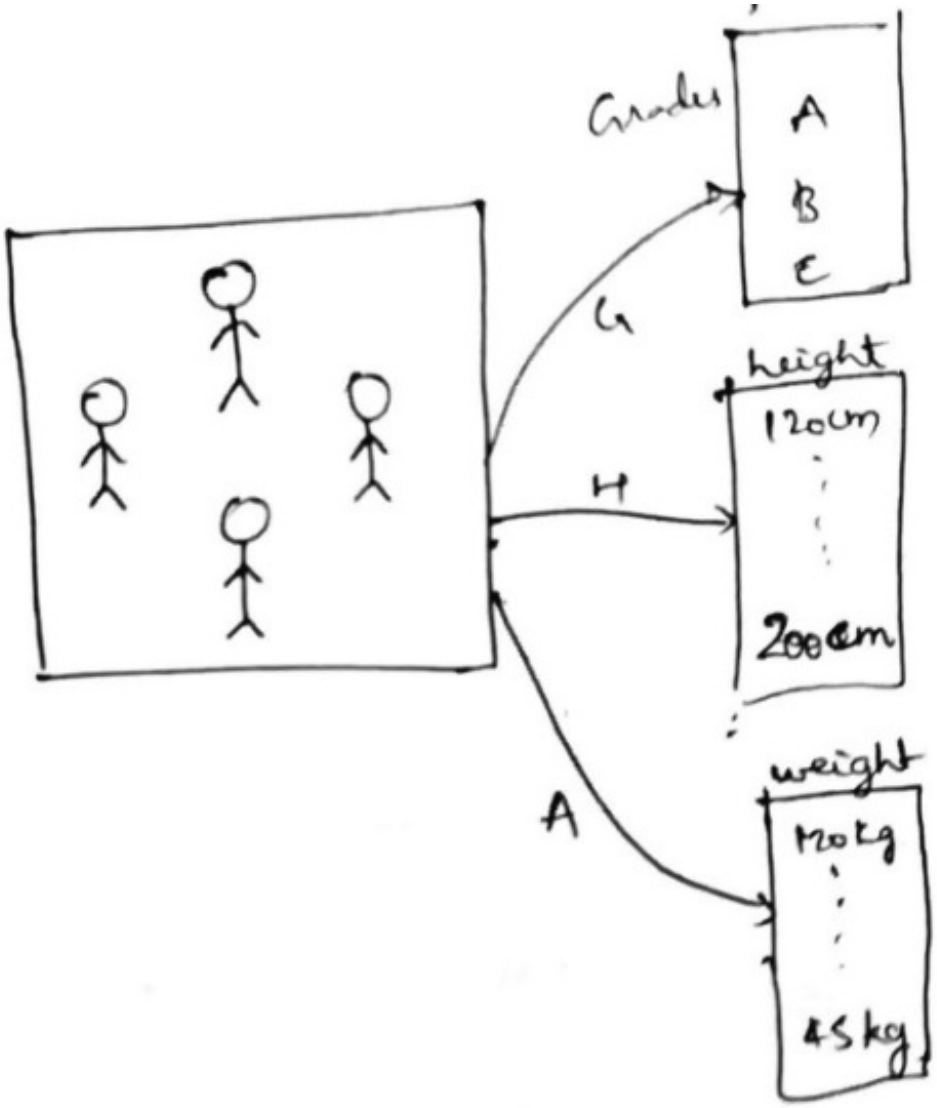


What random variable variable says is that it takes all elements i.e., each and every element in the set and classify it as **Text or noText(1 or 0).**

Now we can calculate **P(class = 0)**[this is short hand of above applying random variable on image and deciding whether there is text or not]



Probability checking.



A random variable can either take continuous values for example the height and weight like **height ranges from 120 cm to 200 cm and weight ranges from 120 Kg to 45 Kg** or they may take **discrete values** like grades as it ranges on **A or B or C or D.**

In this we mainly focus on discrete random variable eg: **Grades(A to c),Rating(1 to 5), Vowels(a to u).**

Distribution:(Marginal Distribution):

Lets the random variable be A,B,C and the what is the distribution of grades G is given as below table which we call as distribution(Marginal Distribution).

Specify the marginal distribution over G, means Specifying **P(G=g) ∀ g ∈ A,B,C**

We denote marginal distribution compactly by P(G).

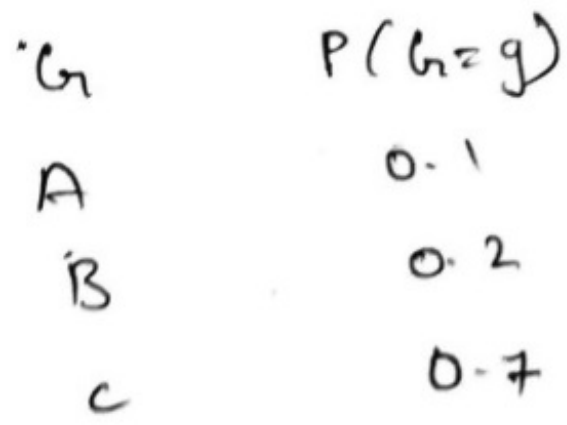
Let take bag full of balls R(Red) G(Green) B(Blue) colours.

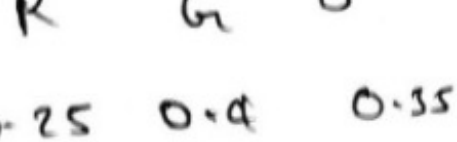
Therefore, from this probability of ball being red is (**Number of red balls /Total number of balls.)**

Else probabilities of R, G, B are 0.25,0.4,0.35

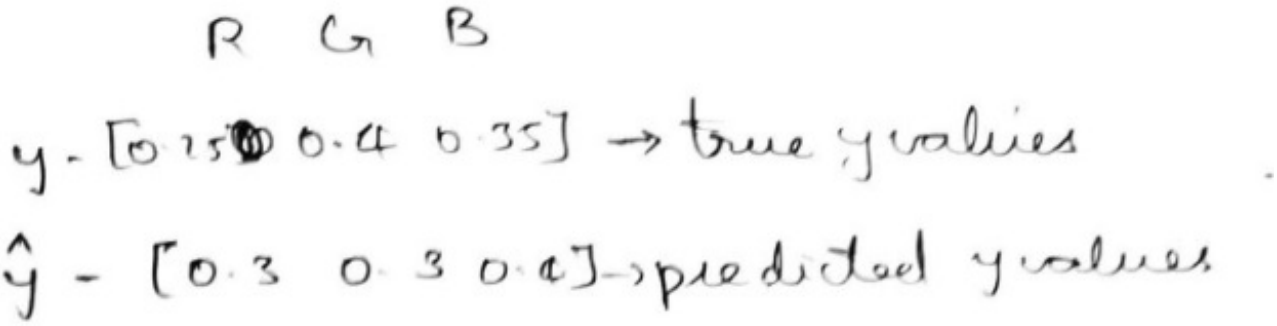
If you allow your friend to peep into the (bag) earn coating balls he will some how guess by estimating them and guess the probability when he is allowed to peep into it once and asked to estimate probabilities of each

Therefore, Tat is the predicted probability y hat.





X	P(X=c)
R	0.25
G	0.4
B	0.35



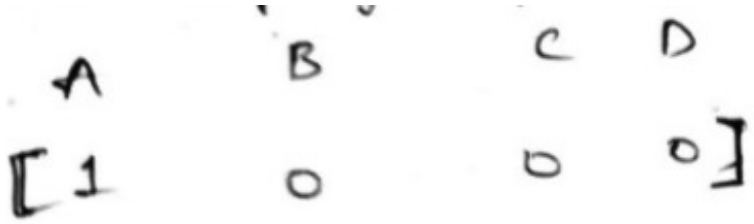
True and Predicted values

We can directly say that he is work in guessing but we wanna say how wrong he was in predicting.

Then you can treat **y and yhat** calculate square error loss.

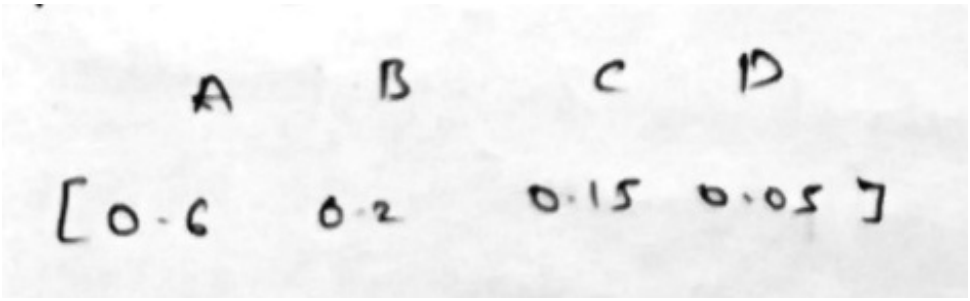
Certain event:

The probability of event if it is sure that it happens then it is called Sure Event and the probability of the event is 1.



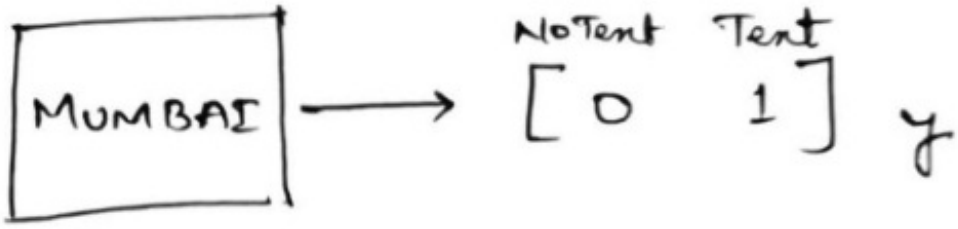
Probability if A winning is certain event

If in match played by A,B,C,D then A has won.



If anyone comes suddenly you have other case where next match you can say I have been Observing the match from 1hr or so the **probability of A winning is 0.6 and that of B is 0.2 , C is 0.15 and D is 0.05 .**

> **Like this we can say in different ways the above are the two different ways of saying.**



In the context of capstone project we will be given an image with text **The basic job is say that It contains Text or not** , the class has only two possibilities whether **text or NoText**.

> What we will do at training time is that we will use sigmoid function we need output to **1 If it contains text and 0 if it doesn't contain text.** We will get the output zero or one like that

Assume 30x30 Image. Therefore, 900 inputs and therefore, 90 bias.

Therefore, if yhat = 0.7 is the thing which we got.

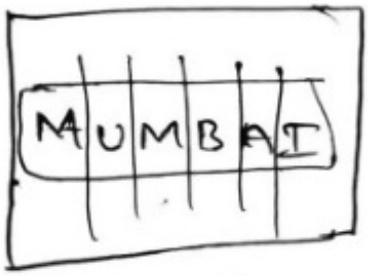
Therefore, yhat = 0.7 is a probability distribution and it is saying the 0.7 probability it contains text and (1–0.7)= 0.3 probability it doesn't contains text.

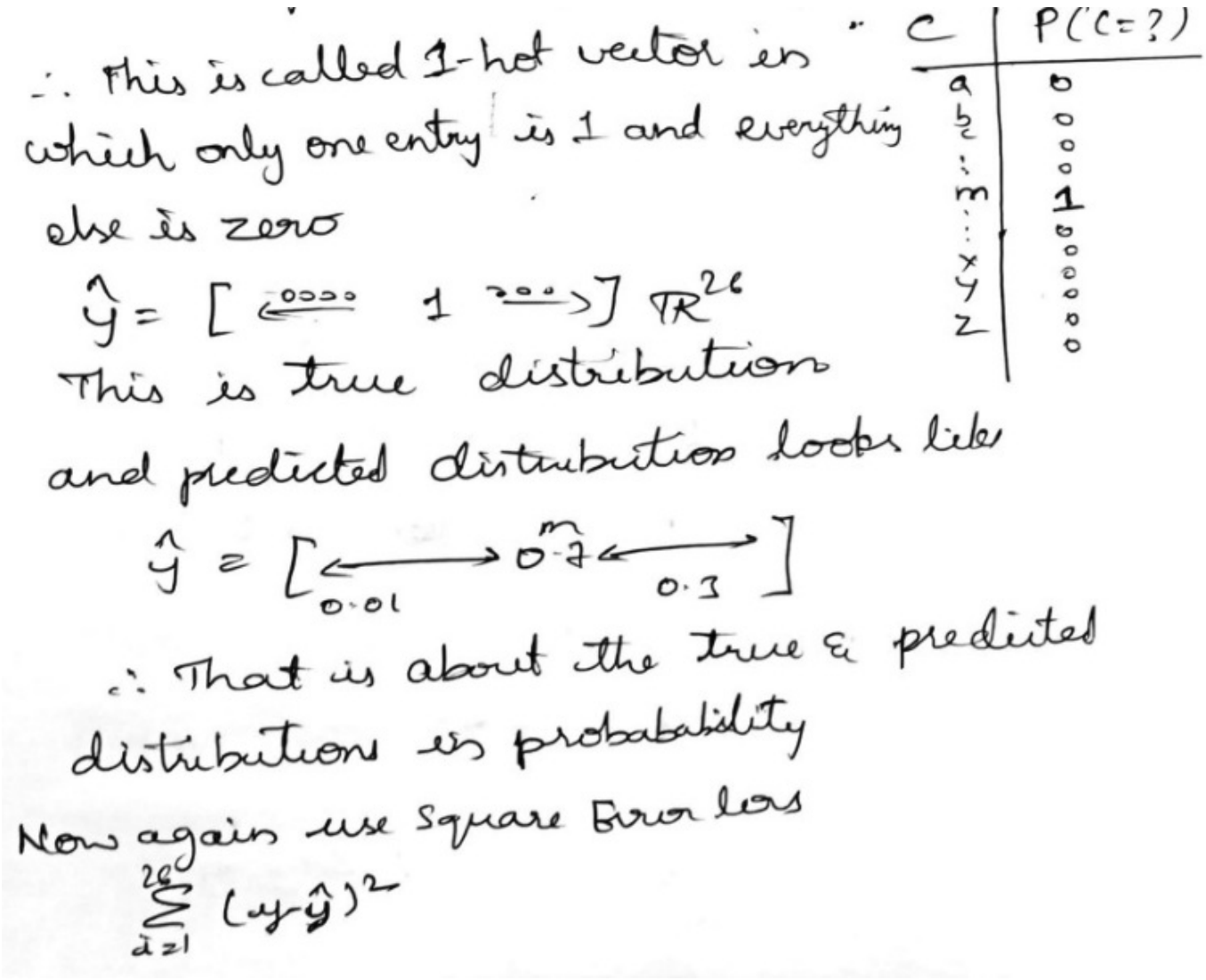
If the model is perfect it **would have returned 1 for having Text** and **0 for not having text.**

Now use square error loss or we will use **Cross Entropy loss which is more grounded for probability.**

This is all for binary classification. Whenever it comes to **multi class classification** is as below

The whole word is segmented into al single letters. AFter seh=gmenting each one is checked with 6 alphabetss in ENglish and comapred





The probability distribution for the Looks like as below for each letter **let the letter be 'm'** .Then all **other letters probabilities will be zero and only 'm' letter probability will be one.**

This is a **1-hot vector** in which **only one entry is 1** and **everything else is zero.**

> **Like this Probability plays a major role in building a Deep Learning Model.**

This is a small try, uploading the notes . I believe in "**Sharing knowledge is that best way of developing skills**".Comments will be appreciated. Even small edits can be suggested.

> **Each Applause will be a great encouragement.**

Do follow my medium for more updates……